Once one or more initial leads have been identified (that is, compounds that show activity against a target and are structurally novel), emphasis is shifted towards exploring more extensively the structure-activity space around those one or more lead molecules. Typically, this is accomplished by selecting and screening compounds that are similar to the initial lead(s). Finally, the accumulated qualitative, and if available quantitative, SAR information is used to optimize the initial leads into preclinical candidates through conventional medicinal chemistry techniques.

Besides similarity and diversity (i.e., dissimilarity), other selection criteria can also be employed. Examples include selecting compounds having desired properties or property distributions as determined by a property prediction algorithm or a quantitative structure-activity model, or exhibiting an optimal fit to a biological receptor as determined by a biomolecular docking algorithm. Compound can also be selected based on 2D and 3D QSAR predictions, and receptor complementarity. Additional details of these and other selection criteria are described in the following patents and patent application, each of which is incorporated herein by reference in its entirety: U.S. Pat. No. 5,463,564, entitled "System and Method of Automatically Generating Chemical Compounds with Desired Properties"; U.S. Pat. No. 5,574,656 entitled "System and Method of Automatically Generating Chemical Compounds with Desired Properties"; U.S. Pat. No. 5,684,711, entitled, "System, Method, and Computer Program for at Least Partially Automatically Generating Chemical Compounds Having Desired Properties"; U.S. Pat. No. 5,901,069, entitled "System, Method, and Computer Program Product for at Least Partially Automatically Generating Chemical Compounds with Desired Properties from a List of Potential Chemical Compounds to Synthesize"; and U.S. Pat. Appl. No.08/963,870, entitled "System, Method and Computer Program Product For Identifying Chemical Compounds Having Desired Properties."

now US 6,421,612

Pat. No. 5,684,711, entitled, "System, Method, and Computer Program for at Least Partially Automatically Generating Chemical Compounds Having Desired Properties"; U.S. Pat. No. 5,901,069, entitled "System, Method, and Computer Program Product for at Least Partially Automatically Generating Chemical Compounds with Desired Properties from a List of Potential Chemical Compounds to Synthesize"; and U.S. Pat. Appl. No. 08/963,870, entitled "System, Method and Computer Program Product For Identifying Chemical Compounds Having Desired Properties." now US 6,421,612

Unlike traditional combinatorial approaches where the entire library is made and tested in a single conceptual step, DirectedDiversity systems physically synthesize, characterize, and test only a portion of that library at a time. The selection of compounds is carried out by computational search engines that combine optimal exploration of molecular diversity with a directed search based on SAR information accumulated from previous iterations of the integrated machinery.

A central task of DirectedDiversity systems is to select an appropriate set of compounds for physical synthesis and biological evaluation. The present invention provides an efficient system and method for selecting such an appropriate set of compounds. That is, the present invention can be used in the systems described in the above listed patents and application, to generate the list of K compounds to be synthesized during each iteration.

5. Structure of Present Invention

It is anticipated that the present invention can be implemented as hardware, firmware, software or any combination thereof, and can be implemented in one or more computer systems and/or other processing systems. In one embodiment, the present invention is implemented by one or more computer systems capable of carrying out the functionality described herein.